

Predikcija vjerojatnosti za zgoditkom u nogometnim utakmicama

Ante Drinovac

Mentor: doc.dr.sc. Marko Đurasević

Sadržaj

- Što je to Expected Goals(xG)?
- Modeli
 - Logistička regresija
 - Slučajne šume
- O bazi
- Implementacija
 - Logistička regresija
 - Slučajne šume
- Rezultati

Što je to Expected Goals(xG)?

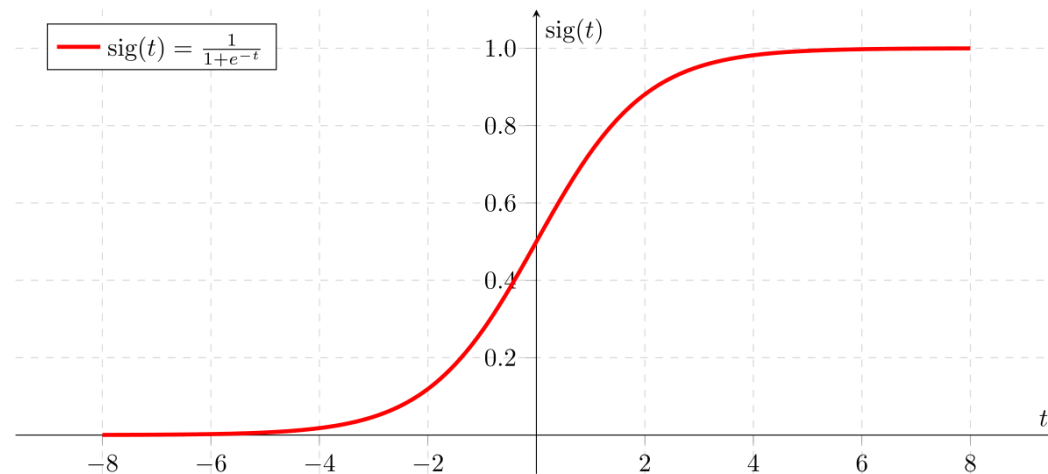
- Vjerojatnost zгодitka pojedinog udarca
- Udarac s xG vrijednosti 0.2 će od 10 udaraca, 2 puta završiti u mreži



Logistička regresija

- Generalizirani linearni model koji predviđa vjerojatnost binarnog ishoda

$$\sigma(\alpha) = \frac{1}{1 + \exp(-\alpha)}$$



Logistička regresija

$$P(1) = \mu \quad P(0) = 1 - \mu$$

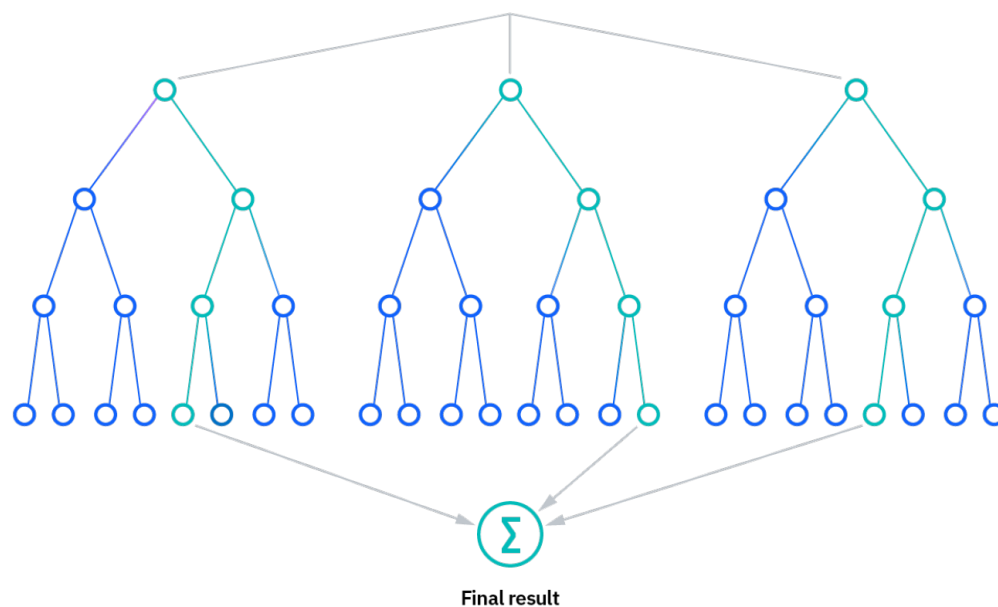
- Pogreška unakrsne entropije:

$$L(y, h(x)) = -y \ln(h(x)) - (1 - y) \ln(1 - h(x))$$

- Gradijentni spust

Slučajne šume

- Složeniji model koji se sastoji od više jednostavnijih stabala odluke



O bazi

- Statsbomb
- Ulazni parametri:
 - under_pressure
 - position
 - play_pattern
 - body_part
 - technique
 - type
 - location
 - shot angle
 - distance



Priprema podataka

- španjolska liga
- 16 sezona
- skup za učenje – 13 sezona
- skup za testiranje – 3 sezone

Logistička regresija

- scikit-learn LogisticRegression
- ugrađene metode:
 - fit(X,Y)
 - predict(X)
 - predict_proba(X)
 - score(X,Y)

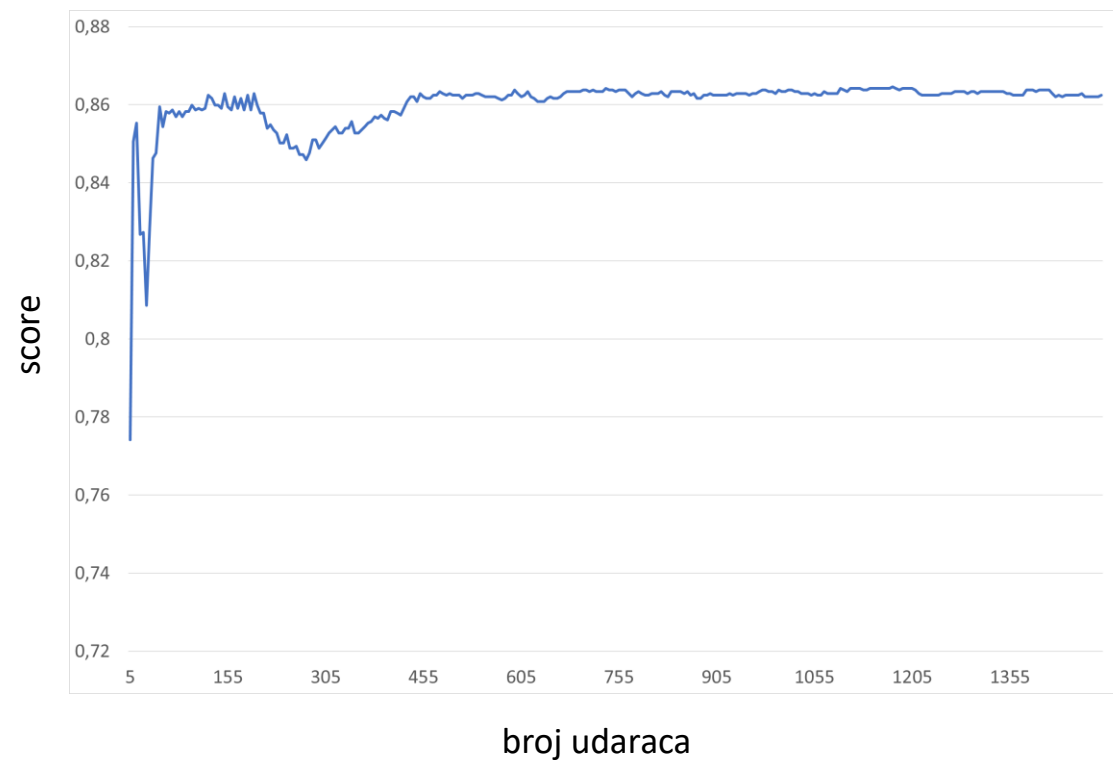
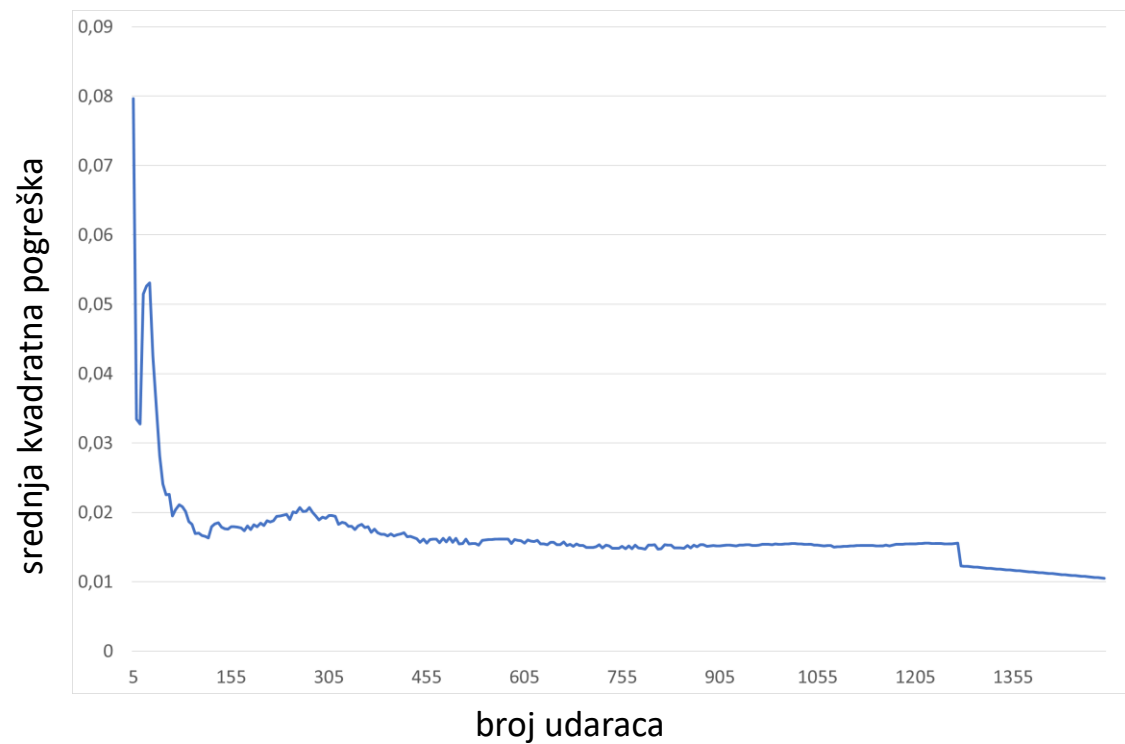
```
def __init__(self):  
    self.model = LogisticRegression(max_iter=1000, solver='saga')
```

Slučajne šume

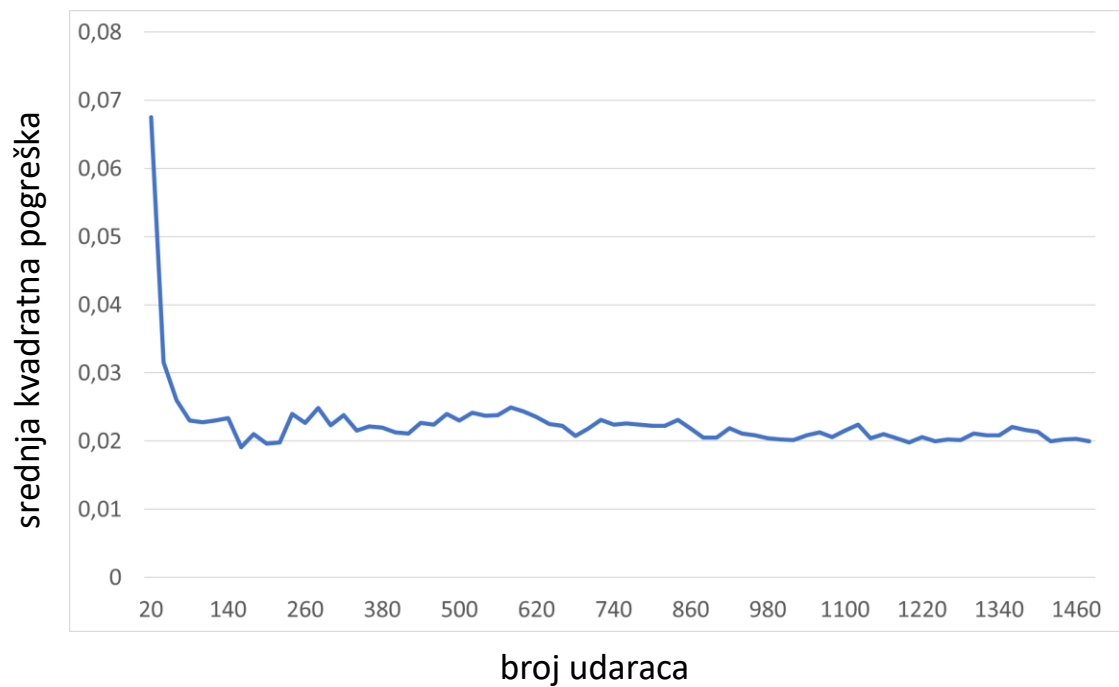
- scikit-learn RandomForestRegressor
- ugrađene metode:
 - fit(X,Y)
 - predict(X)
 - score(X,Y)

```
def __init__(self):  
    self.model = RandomForestRegressor(max_depth=6)
```

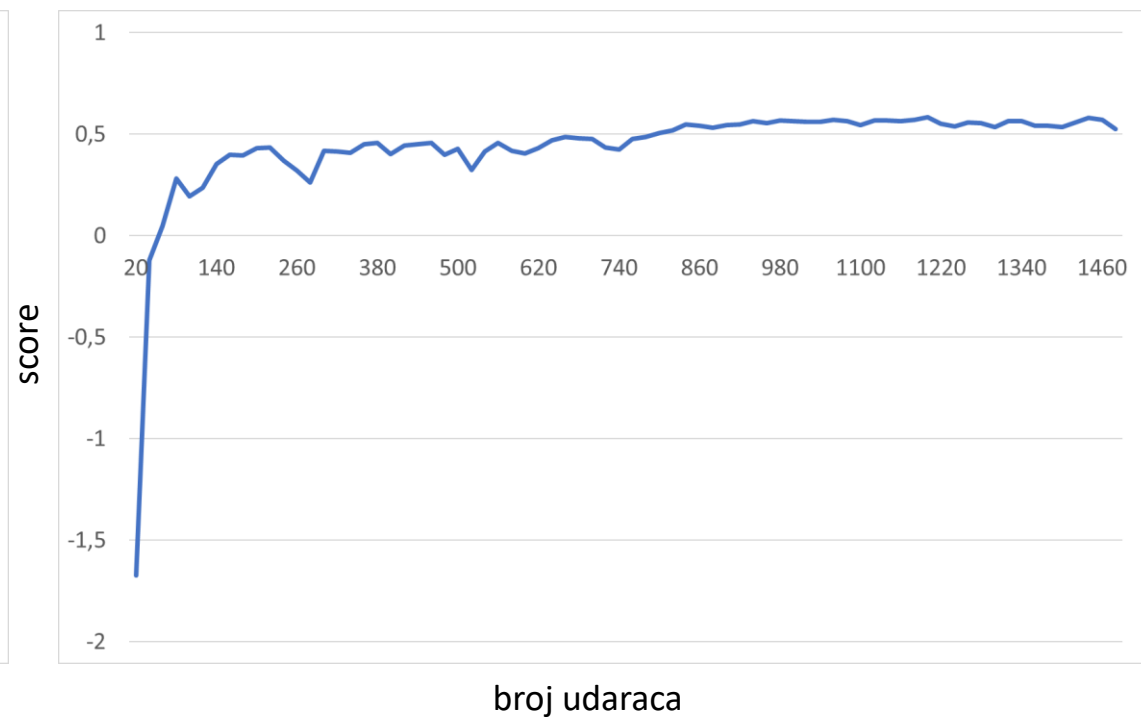
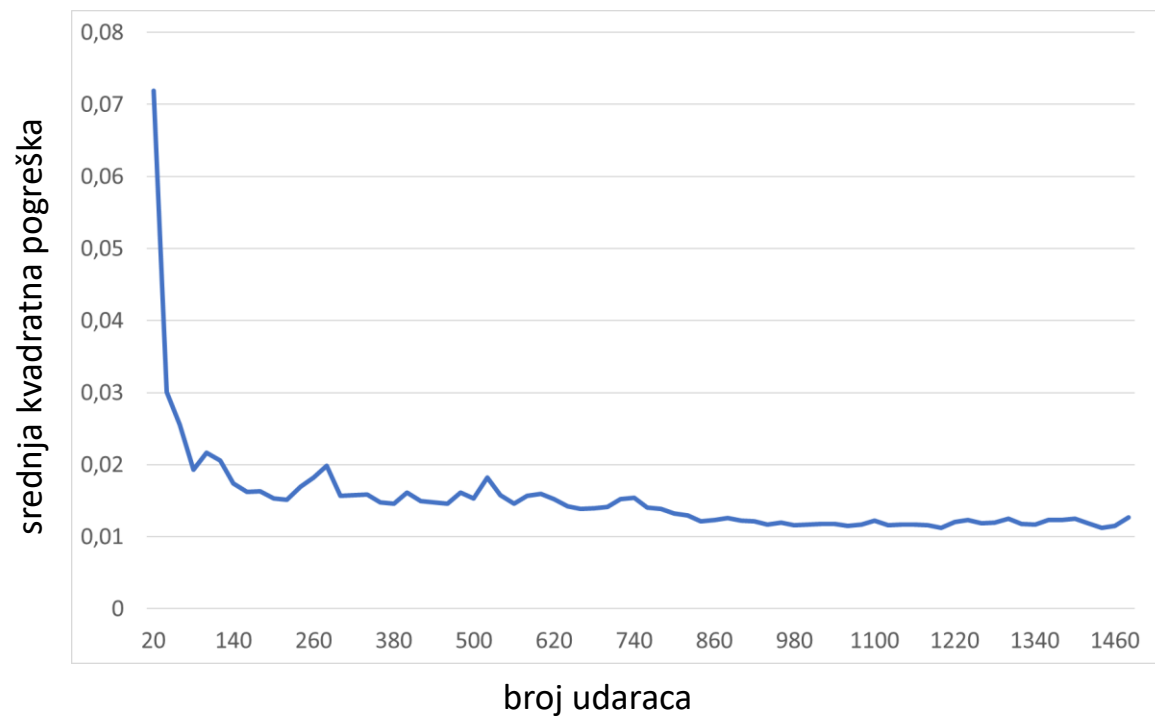
Logistička regresija



Slučajne šume – bez ograničenja dubine



Slučajne šume – s ograničenjem dubine



Usporedba rezultata

	Logistička regresija	Slučajne šume	
		max_depth=0	max_depth=6
score	0.863	0.257	0.694
srednja kvadratna pogreška	0.014	0.020	0.008
stvarna suma xG	304.945	304.945	304.945
predviđena suma xG	333.922	357.674	337.679

Zaključak

- vidljiv pad srednje kvadratne pogreške
- model slučajnih šuma razmjerno uspješniji
- dodatna poboljšanja

Hvala na pažnji